DOCUMENT RESUME

ED 144 947 TM 006 070

AUTHOR Pedulla, Joseph J.; And Others
TITLE Shifts in Teacher Ratings of Students Resulting from Standardized Test Scores.
SPONS AGENCY Carnegie Corp. of New York, N.Y.; National Inst. of Education (DHEW), Washington, D.C.; Russell Sage Foundation, New York, N.Y.; Spencer Foundation, Chicago, Ill.
PUB DATE Apr 77
NOTE 16p.; Paper presented at the Annual Meeting of the American Educational Research Association (61st, New York, New York, April 4-8, 1977); Not available in hard copy due to marginal legibility

EDRS PRICE MF-$0.83 Plus Postage. HC Not Available from EDRS.
DESCRIPTORS Academic Achievement; Elementary School Students; *Elementary School Teachers; *Expectation; *Feedback; Informal Assessment; Mathematics; Primary Education; Reading; *Standardized Tests; Teacher Attitudes; Testing Problems; *Test Results
IDENTIFIERS Ireland; Pygmalion Effect

ABSTRACT

An experimental design was employed to assess the impact of standardized tests on 47 second grade Irish teachers' judgments of their 1,566 students. Teachers receiving test results showed greater shifts in their ratings of students than teachers not receiving test results. However, these shifts tended to be in the positive direction and for only a small number of students. Results indicated that standardized test results did alter second grade teachers' expectations for their students in a small percentage of cases (10% or less). This study also indicated that teachers tended to raise, but not lower, their ratings of students' performance as a result of receiving test results. For at least 90% of the cases, however, the information provided by standardized test results seemed to either corroborate the teachers' existing expectations or to be too weak to alter their existing expectations. Since this study took place in Ireland, the relevance of these findings for the United States was also discussed. The authors cited reports indicating that Irish teachers have more favorable attitudes toward testing than American teachers. Therefore, they concluded that the influence of test results in American teachers' expectations would be even less than the influence on Irish teachers described in this study. (GDC)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED144947

Shifts in Teacher Ratings of Students Resulting from Standardized Test Scores[1]

Joseph J. Pedulla, Boston College
Peter W. Airasian, Boston College
George F. Madaus, Boston College
Thomas Kellaghan, St. Patrick's College, Dublin, Ireland

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

One area of concern in the current controversy over the effects of standardized testing is the impact of standardized test results on teachers' expectations for their students. The recent resurgence of interest in this issue was provided by Rosenthal and Jacobson (1968), who argued that providing teachers with contrived and inflated test results increased students' subsequent test performance. More precisely, the reasoning went as follows: (1) teachers have expectations for their students, (2) teachers attend to the test results in judging pupils, (3) by providing contrived test results to teachers, their existing expectations for students can be altered, (4) the change in expectations will be translated into changes in the teacher-student interaction process, (5) these changes in interactions will reflect the new expectations for pupils which were based on their contrived test scores, and (6) the pupils will conform to the new behavior expected of them.

Rosenthal and Jacobson's results were widely publicized and, to this day, are accepted by many people in education despite the fact that the study was strongly criticized for its methodological and analytical flaws (Elashoff and Snow, 1971; Snow, 1969; Thorndike, 1968). The results of many replication attempts in the

---

[1]Paper presented at the annual meeting of the American Educational Research Association, New York, April, 1977. Work on this study was supported by funds from the Carnegie Corporation of New York, the Russell Sage Foundation, the Spencer Foundation, and the National Institute for Education. The statements made and the views expressed are solely the responsibility of the authors.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Joseph J. Pedulla
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

TM006 070

regular classroom setting failed to corroborate the Rosenthal and Jacobson (1968) findings (cf. Fleming and Anttonen, 1971; Jose and Cody, 1971; Mendels and Flanders, 1973; Wilkins and Glock, 1973). Thus, studies attempting to artificially induce raised expectations in regular classroom teachers by using contrived test score data have generally failed to do so.

There are many possible explanations for the failure of induced expectancy studies to find strong effects. One is that teachers have well-formed, relatively stable expectations for their students based on classroom observation and past teachers' comments (Airasian, Kellaghan and Madaus, 1975; Rist, 1970; Willis, 1972). A single, additional piece of information, such as a standardized test score, may in fact have very little influence on a teacher's pre-existing expectation for a student. This may be particularly true when the test results are false, since the results may be too discrepant from the teacher's existing expectations for the student to be accepted (Fleming and Anttonen, 1971; Pedulla, 1976).

Another possible explanation for the inability of induced expectancy studies to find effects may be due to the fact that only raised expectations were induced. It seems reasonable to assume that, if standardized test results affect expectations, they may lower, as well as raise, some expectations. By examining only raised expectations, researchers are looking at only half of the phenomenon and thus may decrease their chances of finding effects.

A third possible explanation for the lack of significant results in induced expectancy studies may be related to the pervasiveness of standardized testing in the United States where these induced expectancy studies have taken place. Assessing the effect of providing teachers with contrived test results becomes confounded with any previous "real" test results in the students' permanent record and with teachers' prior experiences with and attitudes toward standardized tests. The researchers cannot disentangle the effect of previous exposure to standardized tests from the effects they are investigating. Thus, the treatment is confounded in

induced expectancy studies, and this confounded treatment may mediate against finding the anticipated results.

The purpose of the present study was to examine the effect of actual standardized test results on teachers' expectations for students as evidenced by changes in teachers' ratings of each student's academic ability. Thus the present study has eliminated many of the problems which may operate against finding effects on teachers' expectations due to standardized test results. Students' actual test scores (not artificial ones) were reported back to the teachers. In this way, the artificial nature of the test results used in previous expectancy studies was eliminated and the influence of standardized test results on lowering, as well as raising, teachers' expectations could be determined. The confounding of previous experiences with and attitudes toward standardized testing with the treatment was minimized by conducting the study in a setting in which teachers and students had virtually no prior experience with standardized testing.

Method

Subjects

A sample of 47 second grade teachers teaching 1566 students was randomly selected from a larger sample of second grade teachers and students participating in a societal experiment designed to investigate a wide range of issues associated with the effects of standardized testing. The larger societal experiment is being conducted in the Republic of Ireland with a nationwide random sample of 230 Irish schools, over 1500 teachers, and over 40,000 pupils at all grade levels. Since teachers and students in the Republic of Ireland had virtually no experience with standardized testing prior to the controlled introduction of such tests in this societal experiment, it was possible to examine the effects of tests on teachers with few confounding conditions.

The second grade was selected for study because most previous expectancy studies have concentrated on the lower grades. It was at the lower grades that Rosenthal and Jacobson found their strongest effects. Further, it can be argued convincingly that it is at the lower grades that teachers' expectations for pupils would be most malleable; by the time a pupil is in the higher grades, the informal communication network operating in most schools, in conjunction with the pupil's history of academic performance, may have essentially "pegged" the pupil with a fairly rigid expectation. The effect of standardized test results on these expectations would, therefore, probably be slight. Since grade two is the lowest grade sampled in the larger study, it was selected for investigation.

Procedure

The standardized norm-referenced tests used in this study were the Drumcondra Test Series mathematics computation and English reading subtests. These tests were built specifically for grade two Irish children and normed using a national random sample of second grade Irish students. In most respects, the Drumcondra Tests were very similar to their United States counterparts.

This study utilized a classical pretest-posttest experimental design with the manipulated variable being whether or not teachers received norm-referenced standardized achievement test results for the students in their class. In the larger experiment schools were randomly assigned to treatment groups. One treatment group did not administer any standardized tests. A second treatment group administered the standardized tests but did not receive the results from these tests. A third treatment group administered and received the results from the standardized tests in the form of raw scores and percentile and standard scores based on the

national norm group. There were 11 teachers and 411 students in the no testing group, 8 teachers and 269 students in the testing but no results group, and 27 teachers and 878 students in the testing with results group.

The standardized tests of English and mathematics were administered in November. A few weeks prior to the test administration, teachers in all three treatment groups provided ratings of each student's position relative to the other pupils in the class in English reading and mathematics computation. Ratings were on a five-point scale with '1' signifying that the student was in the bottom fifth of the class and '5' signifying that the student was in the top fifth of the class. In late January or early February, test results, in the form of raw scores, percentiles and standard scores, were reported back to those teachers who received results. In May, all teachers rerated each of their students in mathematics and English.

Changes in the pupil ratings were examined for the three treatment groups to determine whether the availability of test information resulted in greater rating changes. The magnitude and direction of the rating changes were examined separately to determine not only whether test results affected teachers' ratings of their pupils, but also whether test results tended to change ratings in a particular direction.

To this end, for each student, three indices were calculated for each of the two rating areas. The first index measured the magnitude of the change in a pupil's rating pair by taking the absolute value of the difference in that pair. For example, if the teacher rated a student to be in the top fifth of the class in mathematics in the fall (a rating of '5') and the middle fifth of the class in the spring (a rating of '3'), the value of the magnitude measure for mathematics for that student would be 2.

The remaining indices dealt with the direction of rating shifts. One index indicated positive rating shifts. If the pupil's final rating was higher than his initial rating, this index was coded '1'; otherwise this index was coded '0'. An index which indicated negative ratings shifts was computed in a similar fashion, i.e., if a pupil's final rating was lower than his initial rating, this index was coded '1'; it was coded '0' otherwise. It should be noted that the means for the variables measuring direction of rating shifts are directly interpretable as the proportion of ratings shifting in that direction.

In sum, each student had a total of six indices, three for mathematics and an analogous three for reading: (1) the magnitude of the change from initial to final rating, (2) whether the final rating was higher than the initial rating; and (3) whether the final rating was lower than the initial rating. Each of these six indices was used as a dependent variable in a univariate, one-factor analysis of variance. In each analysis, the independent variable was treatment group membership, i.e., no testing, testing but no results, or testing with results. Post hoc comparisons of means were conducted when statistically significant differences were found.

Results

Similar results were obtained for the corresponding rating-change measures for English and mathematics. This similarity was partially a function of the high correlations between magnitude, positive or negative rating shift measures in mathematics with the corresponding measure in English. The correlation matrix is presented in Table 1 and shows that these correlations were approximately .60. Thus, there was a "halo" effect in the ratings, i.e., a change in the English ratings for a student tended to be accompanied by a similar change in the mathematics ratings.

- - - - - - - - - - -

Insert Table 1 here

- - - - - - - - - - -

The positive correlations between the magnitude variables and the direction variables in the same subject area are partially a function of the manner in which these variables were constructed. Non-zero change, in terms of magnitude must, by definition, be related to positive and negative rating shifts. Similarly, the negative correlations between variables measuring positive rating shifts and variables negative rating shifts in the same subject area are a function of the manner in which the variables were constructed, since ratings can shift in one and only one direction. The correlations are of a high enough magnitude to indicate that the analysis of variance results are dependent and must be interpreted in light of this dependency.

The means and standard deviations for the variables are presented in Table 2. The analysis of variance results for all six dependent measures are presented in Table 3. The analysis of variance results for both the English and mathematics magnitude of rating shift variables were statistically significant ($p < .05$). Post hoc comparisons of means, using the Tukey test for honestly significant differences (Winer, 1971), showed that the group which received test results exhibited significantly greater magnitude of rating change than: (1) either of the other two groups for the English ratings and (2) the no testing group for the mathematics ratings.

- - - - - - - - - - - - - - -

Insert Tables 2 and 3 here

- - - - - - - - - - - - - - -

Statistically significant differences were also found for both the English and mathematics variables measuring positive rating shifts (cf. Table 3). That is, for both English and mathematics, there was a difference between the three treatment groups in the proportion of ratings which were raised. Post hoc comparisons of

of means showed that the group which received test results raised more of their ratings in English than the group which tested but received no results. In mathematics, the group which received test results raised more of their ratings than the group which did not test.

The means for the positive rating change measures are directly interpretable as the proportion of ratings shifting upward. Keeping this interpretation in mind, one can see from Table 2 that between 13% and 23% of all the ratings shifted upward. By using the mean from either the no testing or the testing but no results groups as a baseline for rating shifts due to factors other chan receiving standardized test results, one can obtain an indication of the proportion of shifts directly attributable to receiving standardized test results. Whether to use the mean of the no testing or testing but no results groups was established by which one was significantly different from the mean of the group which received results. This approach yields the most liberal estimate of the proportion of rating changes directly attributable to receiving standardized test results. Using the appropriate baseline mean, one finds that 7% of the English ratings and 10% of the mathematics ratings shifted upward solely as a consequence of receiving standardized test results.

No significant differences among treatment groups were found for the negative rating shift measures. The means in Table 2 do show, however, that roughly 20% of the ratings in all groups shifted downward. Since roughly 20% of the ratings also shifted upward, approximately 60% of all ratings did not change at all.

Thus, the results from this study show a "halo" effect in the ratings. They also show that less than half of the teachers' ratings of pupils changed at all over the course of the school year. Of those 40% or so of the ratings which did change,

only a small percentage of the changes (7%-10%) could be attributed directly to being in the group which received test results. Finally, the results from this study indicate that receiving standardized test results tended to raise, but not lower, teachers' ratings of their students.

Discussion

Teachers have very strong and stable expectations for their students early in the school year. Over half of these expectations are so fixed that they do not change at all over the course of the school year. Evidence of this rigidity in expectations was provided by the fact that 60% of the initial-final rating pairs for both English and mathematics on a five-point scale were identical. This finding corroborates the results of other studies which indicate that teachers form expectations early in the school year based on many types of cues from or information about the student and that these initial expectations are lasting (Fleming and Anttonen, 1971; Pedulla, 1976; Rist, 1970; Wilkins and Glock, 1973). It should be noted that these expectations were formed independent of any standardized test results.

The changes that do occur in teacher expectations, as evidenced by their ratings of students, showed a "halo" effect. Teachers do not seem to discriminate between English and mathematics very much. If they raise their estimates of a student in English, they tend to raise their estimate in mathematics also. It seems that teacher expectations are formed and altered at a level more global than the specific subject area level. Students are viewed as "smart" or "slow" in general, not "smart" in some areas and "slow" in others.

How, then, do standardized test results fit into the overall expectancy picture? The results from this study indicate that standardized test results do alter grade 2 teachers' expectations for their students in a small percentage

of cases (10% or less). However, contrary to the detrimental effect that some critics of standardized testing have claimed, this study indicated that teachers tended to raise, but not lower, their ratings of students' performance as a result of receiving standardized test results. However, it must be emphasized that in the vast majority of cases, standardized test results had no influence on teachers' ratings. Thus, the new piece of information which standardized test results provide to teachers seems to either corroborate their existing expectations or be too weak to alter existing expectations for at least 90% of the cases.

Since this study took place in the Republic of Ireland, the relevance of the findings for the United States must be addressed. One comparison that can be made is between American and Irish teachers' attitudes and opinions toward standardized tests. The items for this comparison were employed in a study conducted in the United States over ten years ago (Brim, Glass and Goldberg, 1965) and again with Irish teachers (Airasian, Kellaghan and Madaus, 1975). A comparison of the results from the Brim, et al. study to the Irish replication found that Irish teachers had more favorable attitudes toward standardized tests than their American counterparts (Airasian, Kellaghan and Madaus, 1975).

From this finding it can be argued that standardized tests would have more influence on Irish teachers than on American teachers. If this is the case, the influence of standardized test results on American teachers' expectations would be slight indeed. This small influence may account for the inability of induced expectation studies to find effects, since in fact the effects may not be there to find.

This study examined the expectancy phenomenon in the grossest of ways, i.e., at the global treatment level. As has been suggested (Brophy and Good, 1974), a clearer understanding of teachers' expectations can be obtained by examining smaller,

selected subgroups, such as teachers within treatment or students within teachers within treatment. Although initial attempts at this approach with the data for this study failed to produce much by way of identifying particular types of teachers or students for whom standardized test results were most influential (Pedulla, 1976), further research in this vein seems necessary.

In sum, the results from this study indicate that standardized test results affect teachers' ratings for a small number of students, and the influence of the results is to raise, but not lower, ratings. Thus, the criticism that standardized tests work to the detriment of some students by categorizing them as poor achievers in the teachers' eyes not only was not supported by this study, but the reverse occurrence was found.

References

Airasian, P.W., Kellaghan, T., and Madaus, G.F. The consequences of introducing educational testing: A societal experiment. Unpublished manuscript, Boston College, 1975.

Brim, O.G., Jr., Goslin, D.A., Glass, D.C. and Goldberg, I. The use of standardized ability tests in American secondary schools and their impact on students, teachers, and administrators. Technical Report #3. New York: Russell Sage Foundation, 1965.

Brophy, J.E. and Good, T.L. Teacher-student relationships: Causes and consequences. New York: Holt, Rinehart and Winston, 1974.

Elashoff, J.D. and Snow, R.E. (Eds.) Pygmalion reconsidered. Worthington, Ohio: Charles A. Jones, 1971.

Fleming, E.S. and Anttonen, R.G. Teacher expectancy or My Fair Lady. American Educational Research Journal, 1971, 8, 241-252.

Jose, J. and Cody, J.J. Teacher-pupil interaction as it relates to attempted changes in teacher expectancy of academic ability and achievement. American Educational Research Journal, 1971, 8 39-49.

Mendels, G.E. and Flanders, J.P. Teachers' expectations and pupil performance. American Educational Research Journal, 1973, 10, 203-211.

Pedulla, J. J. The influence of standardized test information on teachers' ratings of their students. Unpublished doctoral dissertation, Boston College, 1976.

Rist, R. C. Student social class and teacher expectations: The self-fulfilling prophecy in ghetto education. Harvard Educational Review, 1970, 40, 411-451.

Rosenthal, R., and Jacobson, L. Pygmalion in the classroom: Teacher expectation and pupils' intellectual development. New York: Holt, Rinehart & Winston, 1968.

Snow, R. Unfinished Pygmalion. Contemporary Psychology, 1969, 14, 197-199.

Thorndike, R. L. Review of Pygmalion in the classroom. American Educational Review Journal, 1968, 5, 708-711.

Wilkins, W.E. and Glock, M.D. Teacher expectations and student achievement: A replication and extension. Ithaca, N.Y.: Cornell University, 1973 (ERIC Document Reproduction Service No. 080-567).

Willis, S. Formation of teachers' expectations of students' academic performance. Unpublished doctoral dissertation, The University of Texas at Austin, 1972.

Winer, B. J. Statistical principles in experimental design, 2nd ed., New York: McGraw-Hill, 1971.

Table 1

Correlations Among the Dependent Variables

Variable	2	3	4	5	6
1. Magnitude of change in English ratings	55	60	33	55	31
2. Magnitude of change in mathematics ratings		32	55	30	55
3. Positive shifts in English ratings			60	-25	-21
4. Positive shifts in mathematics ratings				-20	-27
5. Negative shifts in English ratings					62
6. Negative shifts in mathematics ratings					

Note: Decimal points are omitted.

Table 2

Means and Standard Deviations[1] by Treatment Group for All Dependent Measures

Variable	Treatment Group			
	No Testing	Testing - No Results	Testing and Results	P
Magnitude of change in English ratings	.39 (.52)	.39 (.54)	.49 (.61)	
Magnitude of change in mathematics ratings	.38 (.54)	.46 (.55)	.55 (.66)	
Positive shifts in English ratings	.20 (.40)	.16 (.37)	.23 (.42)	
Positive shifts in mathematics ratings	.13 (.34)	.18 (.39)	.23 (.42)	
Negative shifts in English ratings	.16 (.37)	.19 (.40)	.20 (.40)	
Negative shifts in mathematics ratings	.22 (.42)	.22 (.42)	.23 (.42)	

[1]Standard deviations are presented in parentheses.

Table 3

ANOVA Results for All Dependent Measures

Source of Variation	df	Sum of Squares	Mean Square	F-ratio	p-value
		Magnitude of Change in English Ratings			
Between groups	2	4.74	2.37	7.04	.001
Within groups	1515	510.44	0.34		
		Magnitude of Change in Mathematics Ratings			
Between groups	2	7.29	3,64	9.33	.001
Within groups	1463	571.12	0.39		
		Positive Shifts in English Ratings			
Between groups	2	1.10	0.55	3.30	.04
Within groups	1515	252.60	0.17		
		Positive shifts in Mathematics Ratings			
Between groups	2	2.29	1.15	7.33	.001
Within groups	1463	228.52	0.16		
		Negative Shifts in English Ratings			
Between groups	2	0.34	0.17	1.12	.33
Within groups	1515	233.02	0.15		
		Negative Shifts in Mathematics Ratings			
Between groups	2	0.02	0.01	0.07	.43
Within groups	1463	258.42	0.18		